Types of cost function:

- Mean squared Error [M.S.E]
- Mean Absolute Error [M.A.E]
- Root Mean Squared Error [R.M.S.E]

M.S.E :

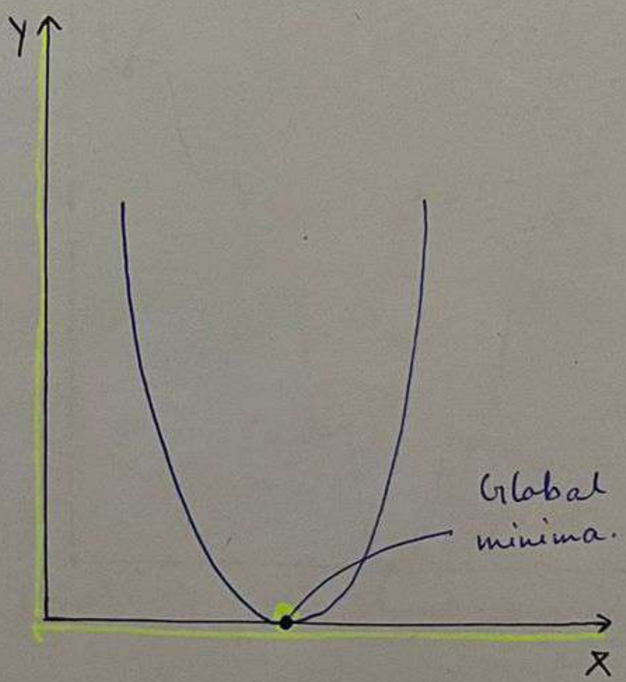
From Mean Squared Error formulae

we have : $\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

where, $\hat{y}_i = \beta_0 + \beta_1 x$

Above equation is similar to quadratic equation : $ax^2 + bx + c$

Curve of quadratic equation is parabola which is same as of M.S.E and also called convex function.



Convex Vs Non- convex function.

Convex function:

In case of convex function when we keep changing the value of β once we will reach the Global minima we have the best fit line. Shown in Fig. 1

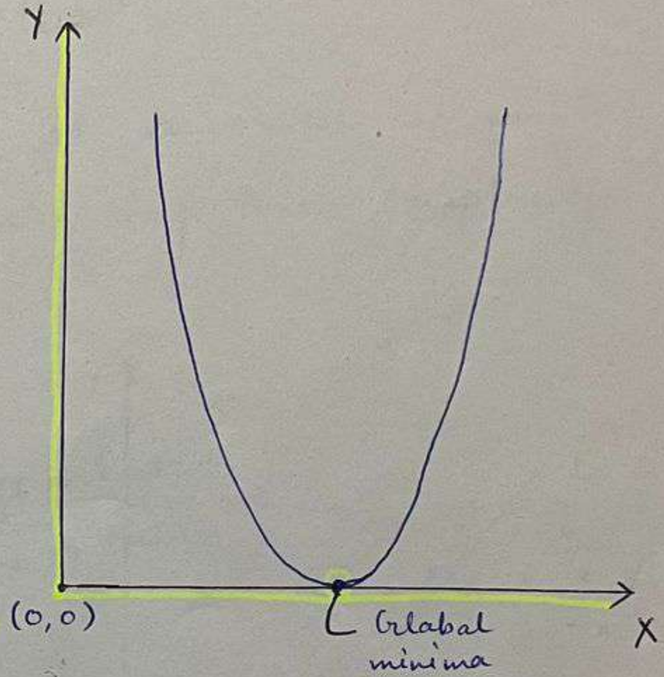


Fig. 1

Non convex function:

In case of non convex function we do have local minima where slop is zero but we can't reach our global minima thus unable to get best fit line for corresponding β value. we have chance to stuck at local minima.

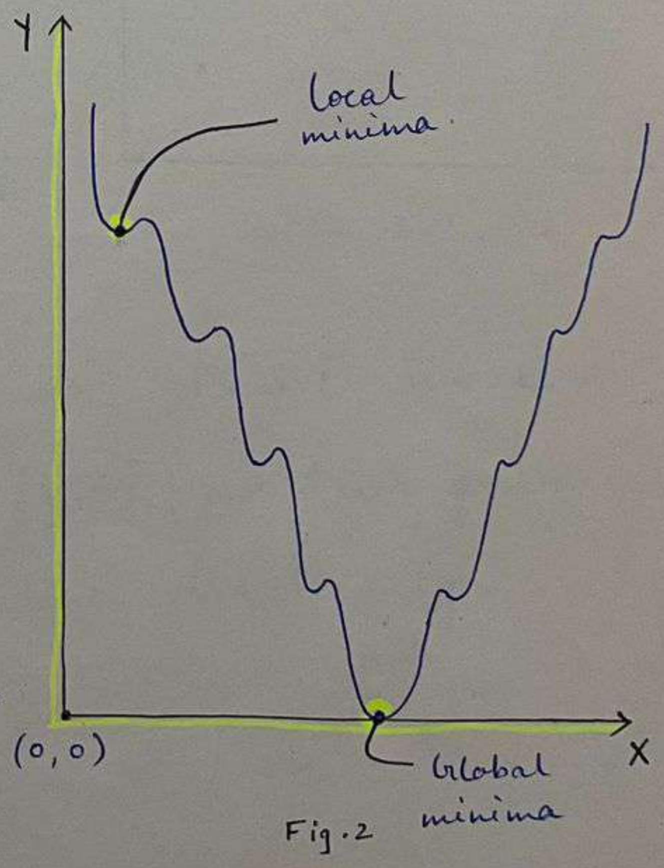


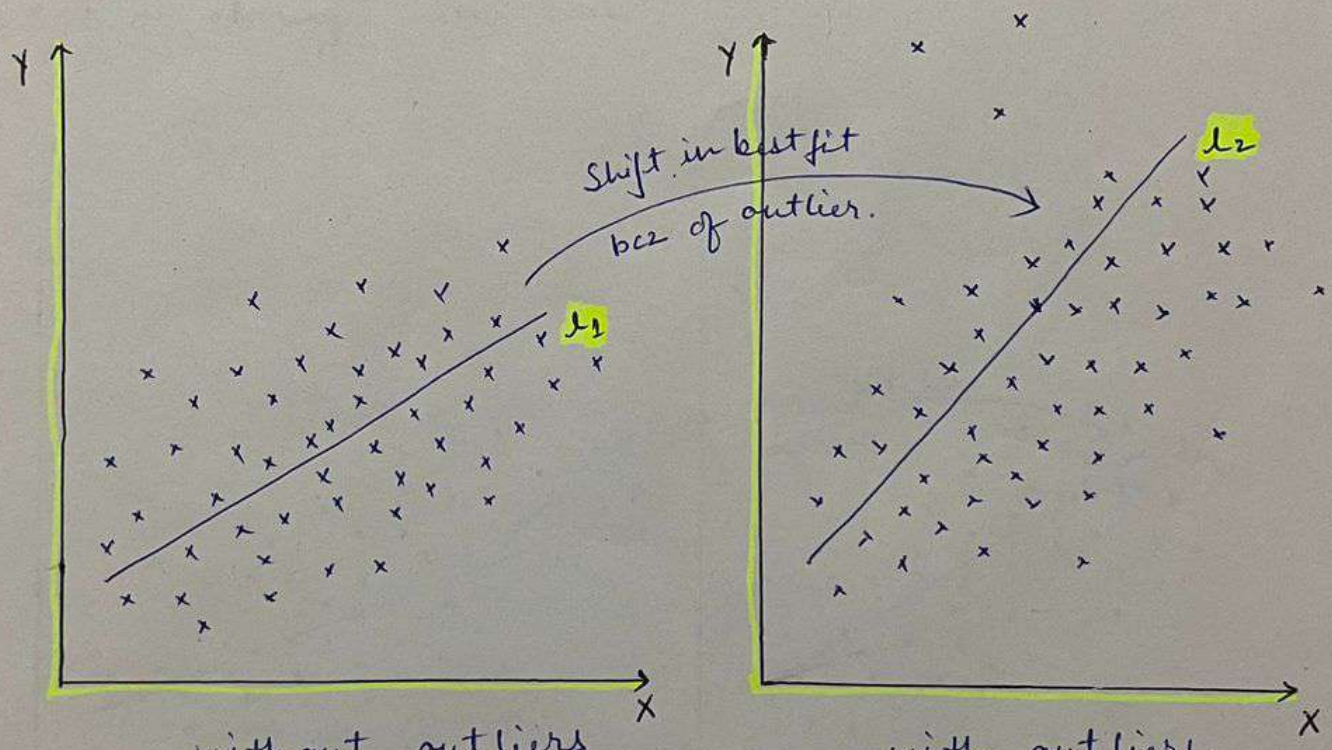
Fig. 2

Advantage of MSE :

- Equation is differentiable
- Only one Global minima & no local minima.

Disadvantage of MSE :

- Not robust to outlier
- In calculation sq. is involved so unit is changed thus increasing time complexity.



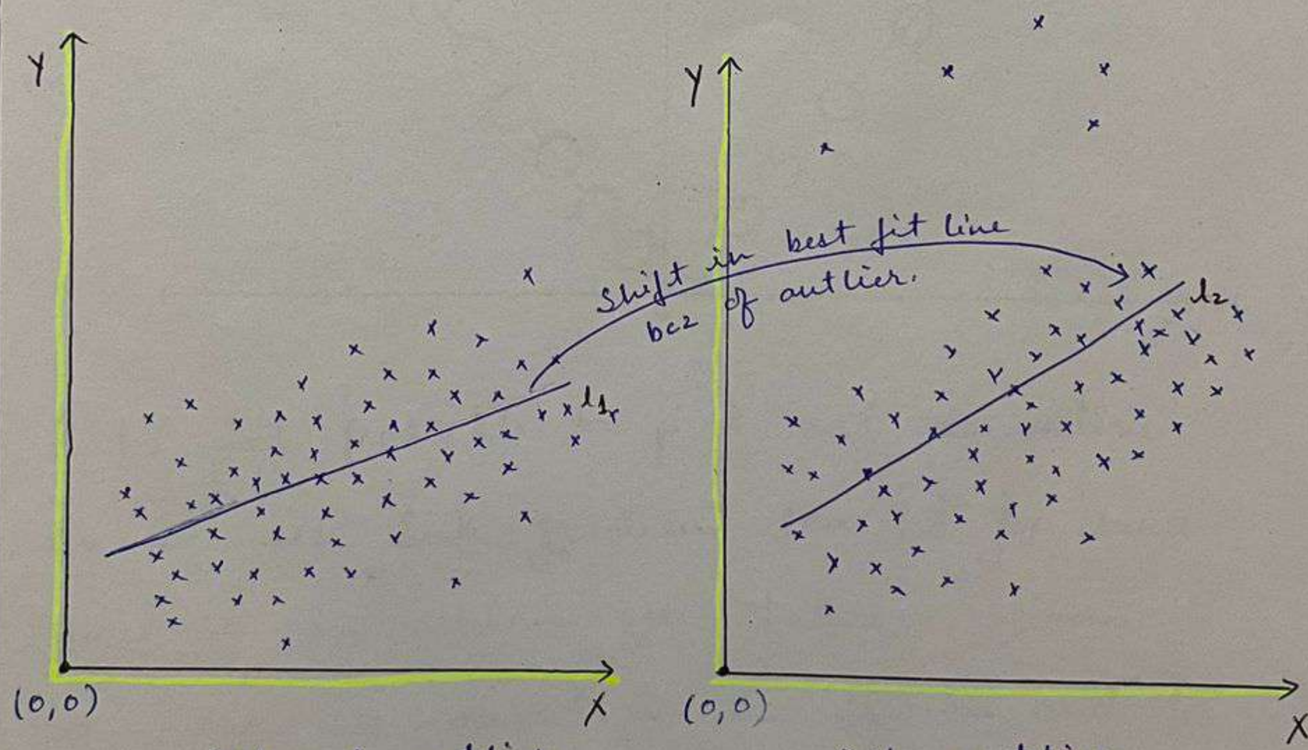
As the calculation in M.S.E involves square in calculation $(y_i - \hat{y}_i)^2$, outlier presence change the best fit line drastically.

MAE :

$$\boxed{= \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \quad \text{Absalute value of } |y_i - \hat{y}_i|}$$

Advantage of MAE :

- Rabust to outlier [as squaring nat invalved]
- No change in Unit.

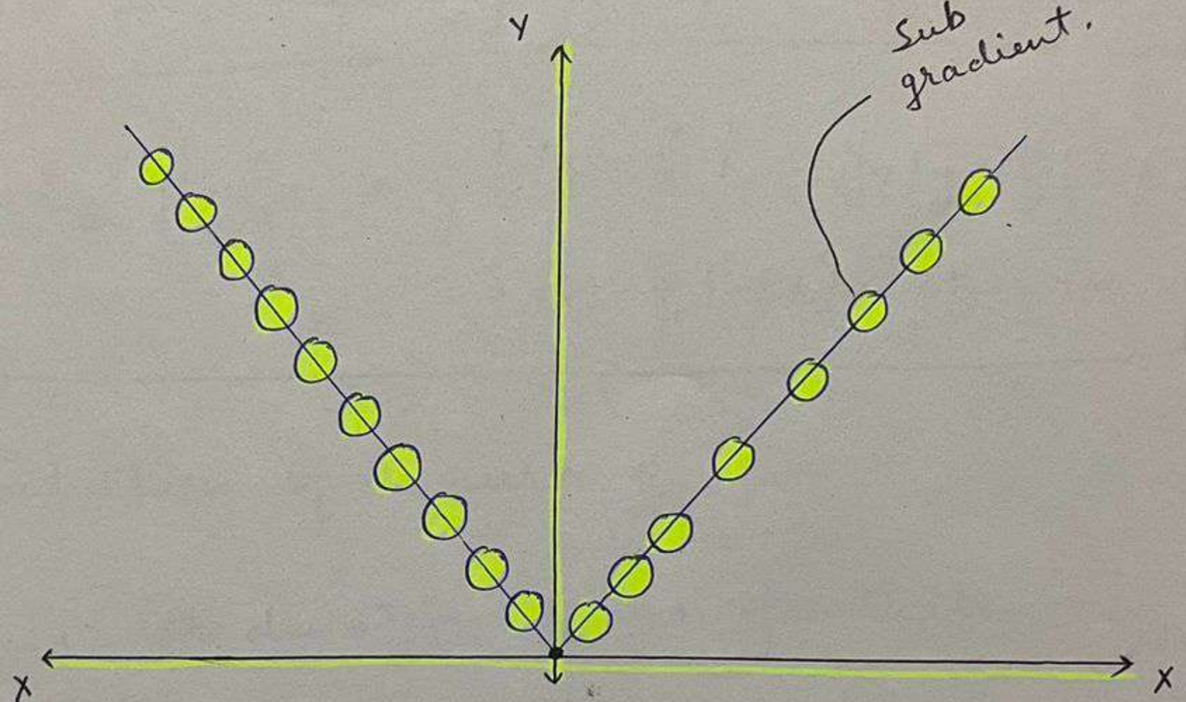


without outlier   with outlier.

Disadvantage :

- Convergence usually takes mare time

optimization in complex task.
As MAE equation is not quadratic thus it forms line. [$|y_i - \hat{y}|$] is involved.



It's not possible to find out derivative. So we use Sub-gradient concept for same purpose in which we takes regions for our consideration as shown and do our calculation. These regions are also called Sub-Gradient.

Huber Loss :

- It combines the best properties of MSE and MAE. It's quadratic for smaller errors and is otherwise linear. It's less ensitive to outlier than MSE. It's used in robust regression.

$$\boxed{L_\delta(y_i, \hat{y}_i) = \begin{cases} \frac{1}{2}(y_i - \hat{y}_i)^2 & , for\ |y_i - \hat{y}_i| \leq \delta \\ \delta \cdot (|y_i - \hat{y}_i| - \frac{1}{2}\delta), & otherwise. \end{cases}}$$

Root Mean square error :

- It's st. deviation of the residuals.

$$\boxed{RMSE = \left[\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \right]}$$